

信用評等如何 統計告訴你

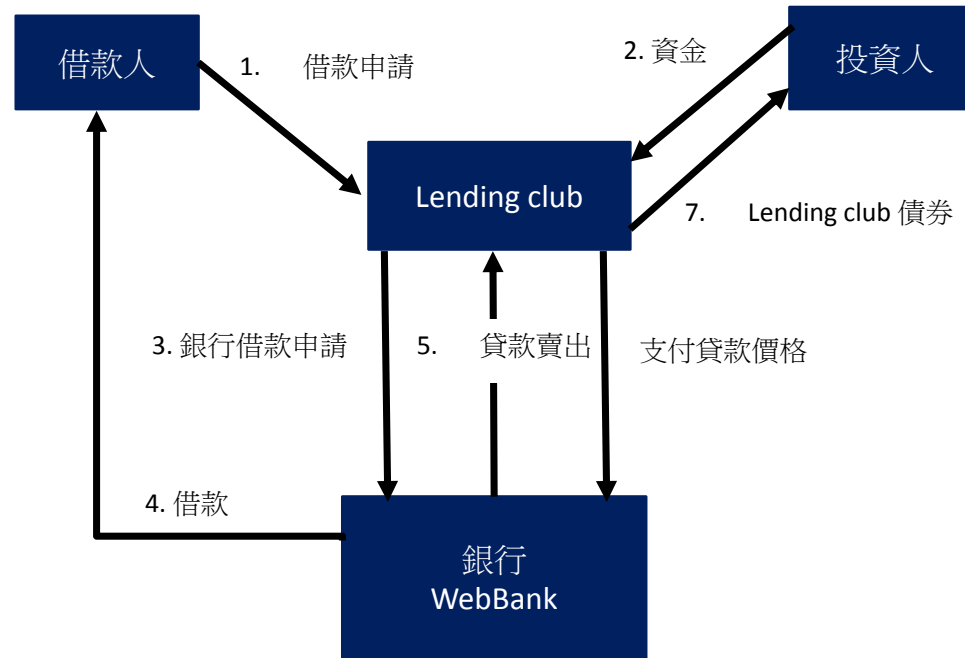
隊名：三峽永和豆漿大王

組員：710433113 粘佑任

710433120 洪允玄

710433104 陳冠豪

Peer to Peer (P2P)



5C分析原則

1. 道德品質(Character)
2. 還款能力(Capacity)
3. 資本實力(Capital)
4. 擔保(Collateral)
5. 經營環境條件(Condition)

信用評等分級



我們的問題

1. 使用不同cluster方法分群並與分級（Grade）真實結果比較，看是否能改進分組方式，但由於無法判定哪個分類方法較佳，所以在最後改成使用cluster找出他們的分類標準。
2. 使用交叉驗證的方法（Cross-validation）找出一個較佳的分類方法可以更快的分級以加速手續的進行時間。
3. 運用縱向分析或是迴歸分析找出可能影響利率的因子，包含美國央行利率、美國公債殖利率

第二章、資料處理

➤ Kaggle平台

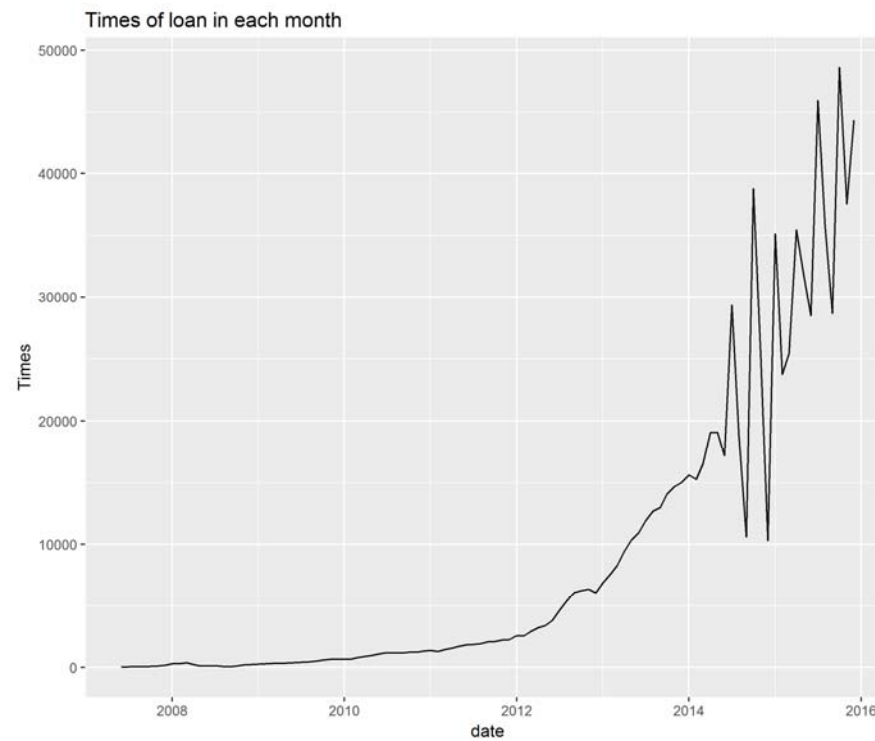
➤ 89萬筆資料分別從2007年蒐集到2015年

➤ 共有74個變數

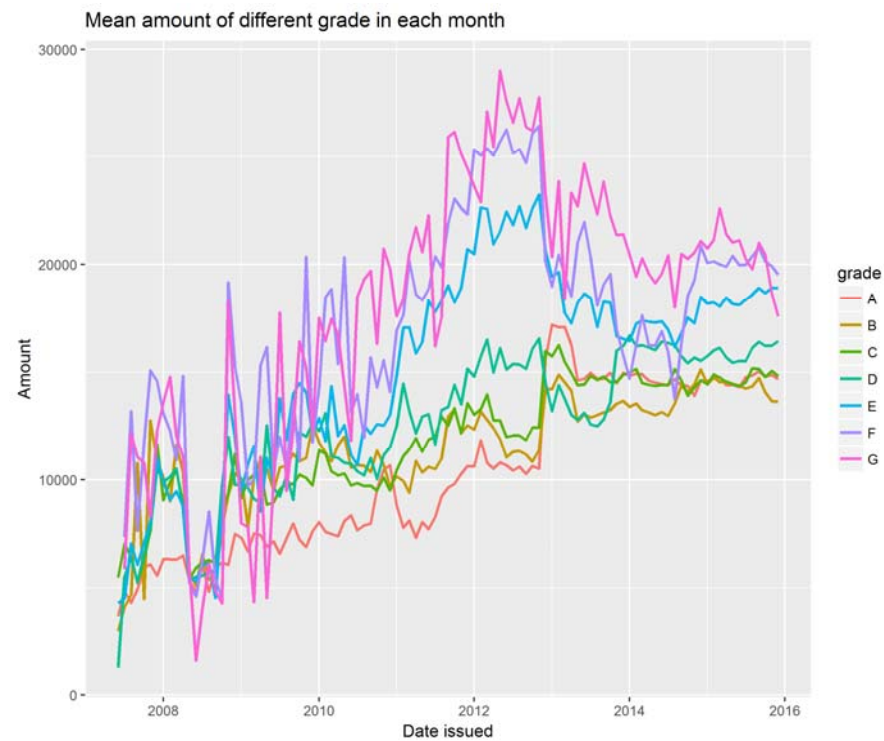
第二章、資料處理

連續型變數		類別型變數	
loan_amnt	申請借款金額	term	貸款支付數量
funded_amnt	承諾借款金額	grade	貸款等級 (7等)
issue_d	貸款資助的		
dti	每月債務支付占收入的比率		
delinq_2yrs	過去兩年逾期超過30天的數量		
revol_bal	總信貸循環餘額		
revol_util	循環限額使用比率		
total_acc	總信用限額		
annual_inc	借款人自填年收入		
tot_cur_bal	借款人在所有帳戶總金額		
tot_coll_amt	借款人收藏品數量		
total_rev_hi_lim	總循環信用佔信用額度的比率		

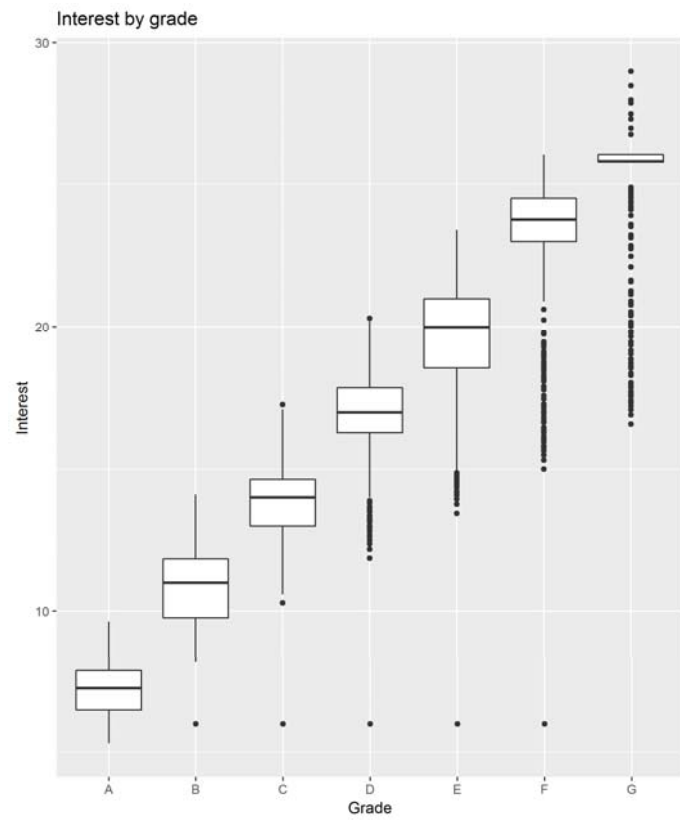
時間與總借貸次數關係



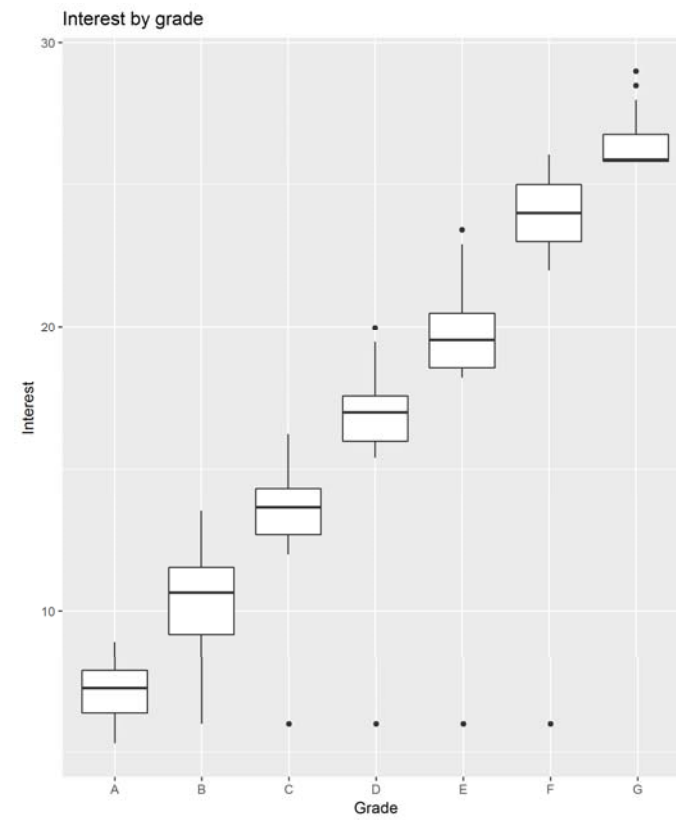
信用評價與申請借貸金額之時間曲線圖



信用評價與借款利率之盒形圖(所有時間)



信用評價與借款利率之盒形圖(2014年至2015年)



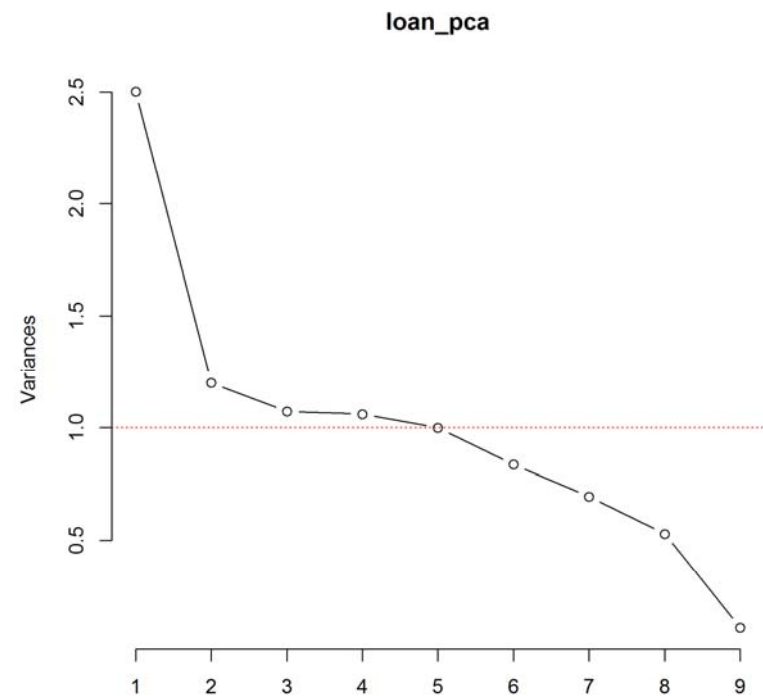
資料分析

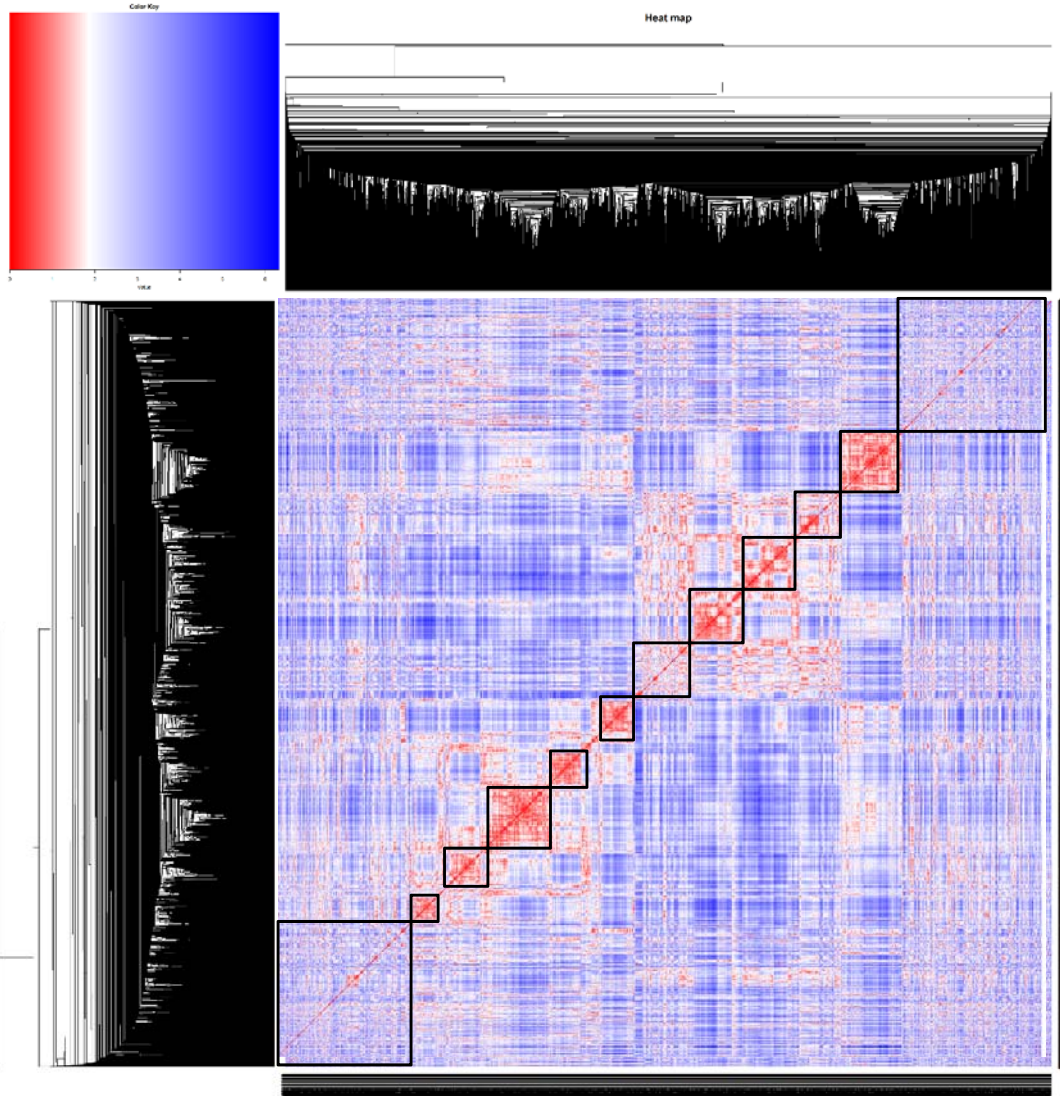
➤主成分分析

➤分群方法(PAM)

主成分分析

	PC1	PC2	PC3	PC4	PC5	PC6
annual_inc	-0.3563	0.117819	0.34703	-0.25164	-0.01625	-0.46031
dti	-0.0304	-0.17924	-0.86254	-0.04126	-0.02603	-0.21343
revol_bal	-0.54047	-0.29341	-0.02985	0.108271	-0.00435	0.333648
revol_util	-0.0565	-0.64473	-0.00419	-0.59498	-0.06765	-0.03881
total_acc	-0.29061	0.482324	-0.32886	-0.0241	0.010358	-0.30251
tot_cur_bal	-0.46012	0.077426	0.089992	-0.21919	-0.01529	-0.25362
tot_coll_amt	0.003295	0.05895	0.008954	0.045183	-0.99657	0.034816
delinq_2yrs	-0.02199	0.455517	-0.12767	-0.62666	0.018799	0.593116
total_rev_hi_lim	-0.52931	-0.05128	-0.04544	0.354805	0.02471	0.347709

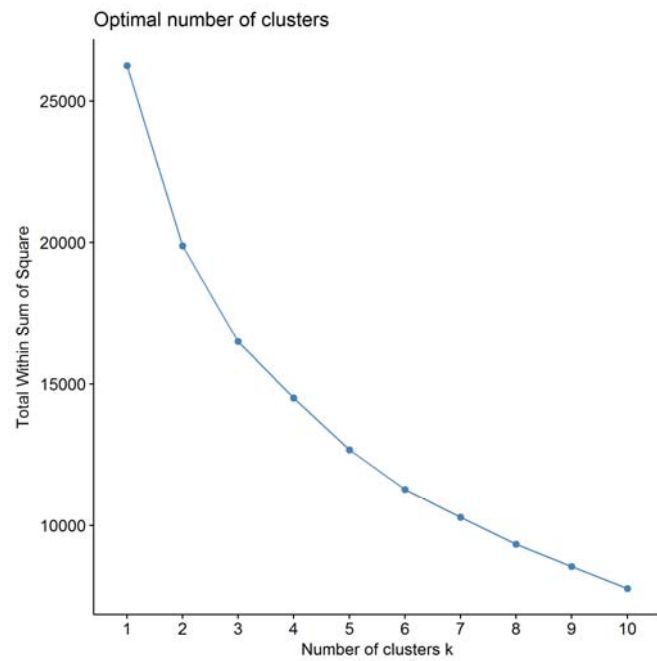




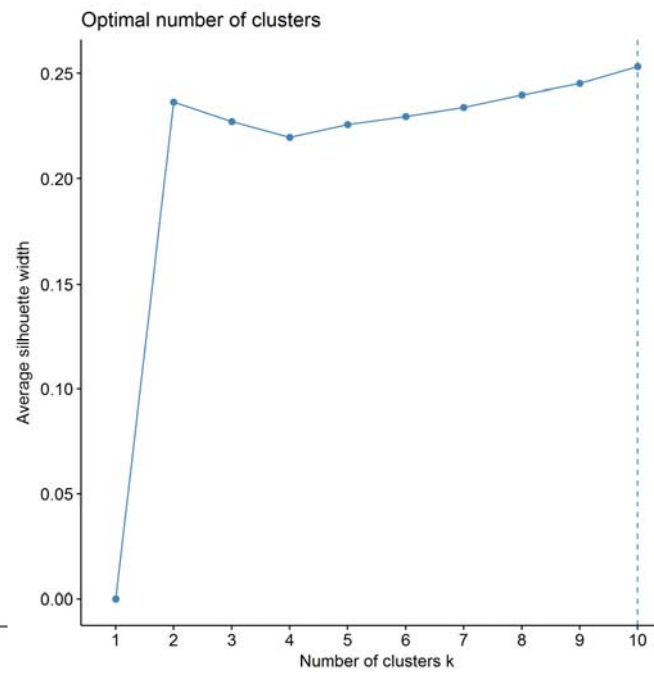
論文引用部分

- 非階層式的分群方法PAM (Partitioning around medoids)
- K-means的演算法來分群
- Park, Jun (2009) 提到使用PAM的方法會比K-means方法還要穩健

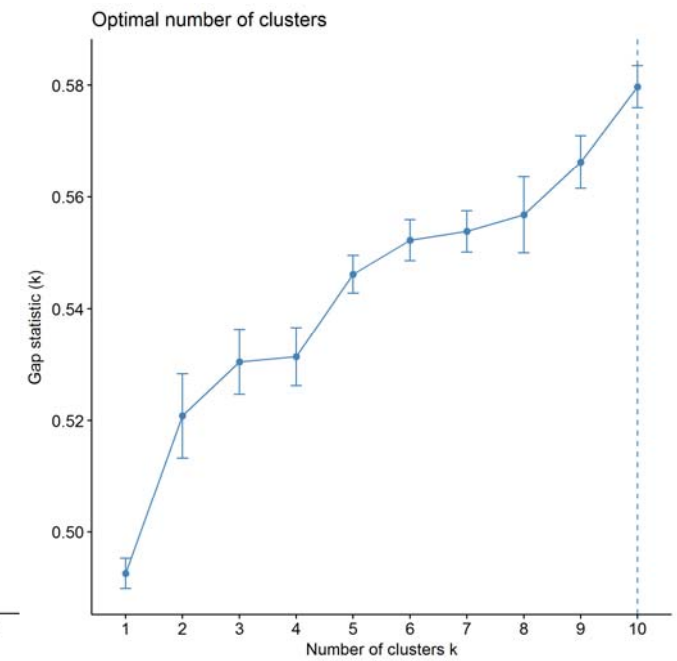
WSS方法視覺化之曲線圖



Silhouette方法視覺化之曲線圖



gap方法視覺化之曲線圖



pam分群與公司信用評的列聯表

	A ^o	B ^o	C ^o	D ^o	E ^o	F ^o	G ^o
1	19716	15233	10517	4926	2476	646	153
2	16782	21110	17508	8364	4342	1266	299
3	10004	21538	23113	13746	7571	2234	543
4	19659	38420	40183	21097	10729	3114	778
5	5221	21075	32309	21612	11582	3349	753
6	19594	27011	24370	12439	6110	1890	501
7	18444	35102	39045	23383	12225	3532	821

	A ^o	B ^o	C ^o	D ^o	E ^o	F ^o	G ^o
1	22725	16844	11215	5015	2510	655	170
2	8460	12441	11294	5847	2943	885	216
3	12297	17634	15929	8362	4424	1243	305
4	14058	29213	31704	16926	8345	2539	646
5	6389	22096	32477	21749	12250	3479	795
6	10686	14309	12363	6237	3113	1012	262
7	10629	18652	20885	12546	7066	2036	441
8	7324	16881	18662	10478	4842	1431	367
9	7285	12261	12598	7040	3413	932	207
10	9567	19158	19918	11367	6129	1819	439

結論

- ▶ 無法找到一個分群方法，與此公司的信用評比一致
- ▶ 信用評價高的人申請借款金額會越低，反之信用評價低的人申請借款金額會越高
- ▶ 資料的ID是沒有重複的情形